



(Cahiers Mathématiques de l'Université de Sherbrooke)

**Titre :** Concepts de dépendance et copules

**Auteur(s) :** Myriam Chabot

**Revue :** CaMUS (Cahiers Mathématiques de l'Université de Sherbrooke)

**Volume :** 4

**Année :** 2013

**Pages :** 48-71

**Éditeur :** Université de Sherbrooke. Département de Mathématiques

**URI :** Repéré à : <http://camus.math.usherbrooke.ca/revue.html>

*Page vide laissée intentionnellement*

# Concepts de dépendance et copules

Myriam Chabot

**RÉSUMÉ** Cet article se veut une introduction à certaines notions de base du concept de dépendance dans le but d'effectuer une entrée en matière avec celui des copules. On y présentera quelques caractéristiques et propriétés fondamentales accompagnées d'exemples pour mieux illustrer les propos expliqués.

## 1 Introduction

Dans une panoplie de domaines de recherche comme la médecine, la finance et la biologie, des chercheurs s'évertuent constamment à comprendre les structures de dépendance entre deux ou plusieurs variables, relations décrites par les fonctions de distribution conjointe. Cependant, déterminer la loi conjointe peut constituer une tâche assez fastidieuse.

Pour l'estimer, certains pourraient être tentés d'utiliser les lois marginales des variables dont il est question. Étant la plupart du temps connues ou supposées connues, celles-ci sont généralement plus facilement mesurables, car une seule dimension doit être considérée : il est plus difficile de faire des hypothèses sur le comportement conjoint de plusieurs variables. En effet, il est possible de s'en servir dans certaines situations. Par exemple, si l'indépendance entre  $n$  variables aléatoires  $X_1, X_2, \dots, X_n$  est assurée, alors la fonction de densité conjointe de ces  $n$  variables s'exprime de la manière suivante :

$$f(x_1, x_2, \dots, x_n) = f_1(x_1) \times f_2(x_2) \times \dots \times f_n(x_n),$$

où  $f_i(\cdot)$  est la fonction de densité univariée de la  $i^e$  variable aléatoire. Par contre, il n'en demeure pas moins qu'établir l'indépendance entre des variables n'est pas nécessairement simple, car, bien qu'elle implique un coefficient de corrélation nul, la réciproque se montre généralement fausse. Une question naturelle se pose : « Quel rôle la dépendance joue-t-elle dans la modélisation statistique ? » Un simple exemple suffit pour se convaincre de l'insuffisance de l'information contenue dans les lois marginales pour résumer le comportement conjoint des

---

J'aimerais remercier M. Taoufik Bouezmarni, professeur à l'Université de Sherbrooke, pour la supervision de cet article rédigé à la suite du cours d'initiation à la recherche suivi en automne 2012. Je voudrais également remercier M. Éric Marchand, professeur à l'Université de Sherbrooke, pour les commentaires apportés lors de la révision de l'article, Félix Camirand-Lemyre, étudiant au doctorat, pour l'aide apportée à la rédaction et la correction de l'article, de même que Charles Fontaine, également étudiant au doctorat, pour l'aide informatique fournie.

variables entre elles. Soient  $X$  et  $Y$  deux variables aléatoires ayant toutes deux comme distribution marginale  $\text{bêta}(2,2)$ . Avec ces hypothèses de départ, on effectue deux simulations de 2000 données chacune et on obtient les deux graphiques présentés à la figure 1.

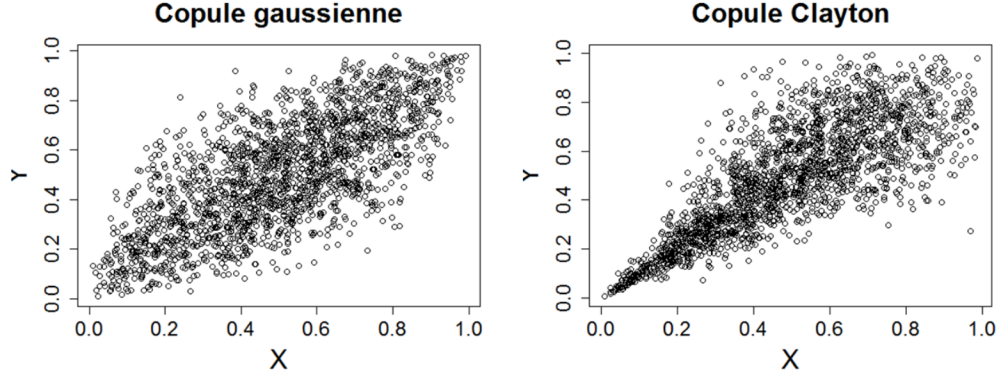


FIGURE 1 – Simulation de données pour comparer le comportement conjoint de deux variables aléatoires ayant chacune comme distribution marginale  $\text{bêta}(2,2)$

On peut alors constater que, malgré le fait que  $X$  et  $Y$  soient de même distribution marginale dans chacune des simulations, la structure de dépendance entre ces deux variables est bien différente. En effet, dans le graphique de gauche, la concentration du nuage de points est plus prononcée pour de grandes valeurs de  $X$ , ce qui confère une meilleure prédiction des valeurs de  $Y$  lorsque l'on connaît de telles valeurs de  $X$ . C'est plutôt le cas pour les petites valeurs de  $X$  dans le graphique de droite. On verra plus tard que la première simulation sera associée à une copule gaussienne, alors que la deuxième sera associée à une copule Clayton.

Les copules constituent un sujet de recherche relativement moderne étant en plein essor depuis principalement les trois dernières décennies. Provenant du nom latin *copula*, qui signifie « *tout ce qui sert à attacher, lien, chaîne* » [Gaf13], la copule est un outil statistique qui modélise la dépendance entre des variables aléatoires, permettant ainsi de considérer de manière distincte la structure de dépendance décrite par la fonction de distribution conjointe et le comportement marginal des variables considérées. En effet, soit un couple de  $n$  variables aléatoires  $(X_1, X_2, \dots, X_n)$  tel que, pour  $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ , la fonction de répartition conjointe est donnée par

$$F(x_1, x_2, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

et la fonction de répartition marginale de la  $i^{\text{e}}$  variable aléatoire, pour  $i \in \{1, 2, \dots, n\}$ , est donnée par

$$F_i(x_i) = \mathbb{P}(X_i \leq x_i).$$

Le célèbre théorème de Sklar [Skl59], qui sera davantage étudié plus loin dans l'article, stipule que si les fonctions  $F_i(\cdot)$  sont continues, alors il existe une unique

fonction de répartition multivariée  $C : [0,1]^n \rightarrow [0,1]$ , que l'on appellera copule, dont les marges sont uniformes et telle que, pour tout  $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ , on puisse écrire

$$F(x_1, x_2, \dots, x_n) = C\{F_1(x_1), F_2(x_2), \dots, F_n(x_n)\}.$$

Autrement dit, il est possible d'associer la loi conjointe multivariée aux lois marginales univariées à l'aide de la fonction copule qui contient l'information nécessaire sur la structure de dépendance des variables aléatoires entre elles.

Bien que le terme *copule* fut utilisé comme objet mathématique pour la première fois en 1959 dans les travaux d'Abe Sklar, mathématicien américain et professeur émérite à l'*Illinois Institute of Technology*, la fonction sous-jacente fut étudiée bien avant, sous d'autres appellations, par exemple dans les travaux de Fréchet [Fré51], de Féron [Fé56] et de plusieurs autres mathématiciens et statisticiens qui s'intéressaient aux fonctions de distribution multivariée ayant des lois marginales fixées [Nel06].

On retrouve notamment plusieurs résultats à la base de la théorie des copules dans des travaux de Wassily Hoeffding dans [Hoe40] et [Hoe41] (voir également [Nel06]). Dans ceux-ci, ce statisticien et probabiliste finnois décrivait des fonctions de distribution standardisée bivariable dont le domaine est le carré  $[-1/2, 1/2] \times [-1/2, 1/2]$  et dont les marges sont uniformes sur l'intervalle  $[-1/2, 1/2]$ . Il travailla également sur les mesures de dépendance invariantes sous transformations strictement croissantes. Malgré l'apport notable de ses travaux, ces derniers ne furent pas immédiatement considérés à leur juste valeur notamment en raison de circonstances historiques (Deuxième Guerre mondiale). D'ailleurs, Fréchet arriva à des conclusions semblables à celles obtenues par Hoeffding malgré le fait qu'elles lui étaient inconnues [Fré51]. De ce fait, aujourd'hui, lorsque l'on fait référence à certains résultats de ce domaine dans la littérature, notamment concernant l'ordre stochastique, il n'est pas rare de trouver ces deux noms juxtaposés. Puis, en 1959, Sklar écrivit un article en utilisant pour la première fois le mot « copule » avec le sens qu'on lui connaît à ce jour (voir [Skl59]), alors qu'il travaillait avec Berthold Schweizer sur la théorie des espaces métriques probabilistes. Pour plus de détails, on réfère le lecteur à [Nel06].

Cet article se veut une introduction à la notion de copules. Pour ce faire, nous commençons par présenter certaines propriétés de base importantes lorsque nous parlons de dépendance entre variables aléatoires. Ceci nous permet ensuite d'explorer les rudiments des copules, pour alors en arriver au théorème de Sklar. Finalement, nous donnons quelques exemples de copules accompagnés de leurs caractéristiques principales.

## 2 Dépendance

Puisqu'il s'agit d'un article davantage dédié aux copules, nous n'explorerons pas de manière exhaustive tous les concepts de la littérature liés à la notion de dépendance. Il importe toutefois d'en présenter l'essentiel pour mieux comprendre

ce que constituent les copules. Pour des fins de simplification et afin de favoriser de meilleures compréhension et visualisation des notions, nous nous restreignons à l'étude du cas bivarié, où seules deux variables aléatoires  $(X_1, X_2)$  sont en relation, bien que la plupart des résultats qui sont présentés se généralisent dans le cas multidimensionnel.

Une notion qui revient tout au long de l'article est celle d'une fonction de répartition bivariée, parfois appelée fonction de distribution cumulative bivariée. Celle-ci peut être définie de deux manières. D'une part, on a

$$F(x_1, x_2) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2). \quad (1)$$

D'autre part, contrairement à cette définition qui fait appel aux variables aléatoires, on peut énoncer des conditions nécessaires et suffisantes pour qu'une fonction soit une fonction de répartition bivariée [Joe01].

**Définition 2.1.** Une fonction  $F : \mathbb{R}^2 \rightarrow [0, 1]$  continue à droite est dite une *fonction de répartition bivariée* si et seulement si :

- 1)  $\lim_{x_i \rightarrow -\infty} F(x_1, x_2) = 0$ , pour  $i \in \{0, 1\}$  ;
- 2)  $\lim_{x_1, x_2 \rightarrow \infty} F(x_1, x_2) = 1$  ;
- 3) (*Inégalité du rectangle*) pour tout  $(\alpha_1, \alpha_2), (\beta_1, \beta_2) \in \mathbb{R}^2$  tels que  $\alpha_1 < \beta_1$  et  $\alpha_2 < \beta_2$ , on a

$$F(\alpha_1, \alpha_2) + F(\beta_1, \beta_2) - F(\alpha_1, \beta_2) - F(\beta_1, \alpha_2) \geq 0. \quad (2)$$

Une fonction quelconque  $F$  satisfaisant l'équation (2) sera dite *2-croissante*. Par contre, malgré ce que peut laisser sous-entendre l'appellation, si  $F$  est croissante par rapport à  $x$  et par rapport à  $y$ , elle ne sera pas nécessairement 2-croissante. La réciproque sera vraie si la condition 1 de la définition précédente est également respectée. Cette dualité entre les approches probabiliste et fonctionnelle de la définition d'une fonction de répartition bivariée est également présente dans la définition de la copule.

*Notation 2.2.* On désignera par  $\mathcal{C}(F_1, F_2)$  la classe des fonctions de répartition bivariées de marginales  $F_1$  et  $F_2$ .

*Remarque 2.3.* Les marginales s'obtiennent de la fonction de distribution :

$$F_1(x_1) = \lim_{x_2 \rightarrow \infty} F(x_1, x_2); \quad (3.a)$$

$$F_2(x_2) = \lim_{x_1 \rightarrow \infty} F(x_1, x_2). \quad (3.b)$$

**Définition 2.4.** La *fonction de survie*  $\bar{F}$  associée à une fonction de répartition cumulative bivariée  $F$  correspond à

$$\bar{F}(x_1, x_2) = \mathbb{P}(X_1 > x_1, X_2 > x_2). \quad (4)$$

Cette dernière peut également s'écrire en fonction de  $F$ ,  $F_1$  et  $F_2$  :

$$\bar{F}(x_1, x_2) = 1 - F_1(x_1) - F_2(x_2) + F(x_1, x_2). \quad (5)$$

## 2.1 Ordre à dépendance positive bivariable

Dans certaines familles paramétriques de distribution multidimensionnelle, nous remarquons que la valeur du ou des paramètres influence la force de la dépendance [Joe01]. En s'inspirant de ce fait, nous pouvons comparer les fonctions de répartition multivariées entre elles et, ainsi, établir un ordre partiel, car relatif aux fonctions comparées, permettant de déterminer si une fonction de répartition cumulative est plus dépendante qu'une autre. La comparaison peut également s'exprimer en termes d'un couple de variables aléatoires.

*Notation 2.5.* Pour parler de convergence en distribution ou en loi, on utilisera le symbole  $\xrightarrow{d}$ .

**Définition 2.6.** Soient  $F, G, H \in \mathcal{C}(F_1, F_2)$  trois fonctions de répartition bivariables. Soient  $(X_1, X_2)$  et  $(X'_1, X'_2)$  deux couples de variables aléatoires telles que  $X_1$  et  $X'_1$  ont la même distribution marginale, tout comme  $X_2$  et  $X'_2$ . Un ordre est dit à *dépendance positive bivariable*, noté  $\prec$ , s'il satisfait les conditions suivantes :

- 1) (*Concordance*)  $F \prec G \implies F(x_1, x_2) \leq G(x_1, x_2)$ , pour tout  $(x_1, x_2) \in \mathbb{R}^2$ ;
- 2) (*Réflexivité*)  $F \prec F$ ;
- 3) (*Symétrie*)  $F \prec G$  et  $G \prec F \implies F = G$ ;
- 4) (*Transitivité*)  $F \prec G$  et  $G \prec H \implies F \prec H$ ;
- 5) (*Bornes*)  $\max\{0, F_1(x_1) + F_2(x_2) - 1\} \prec F \prec \min\{F_1(x_1), F_2(x_2)\}$ ;
- 6) (*Invariance à la limite distributionnelle*)  $F_n \prec G_n$  (pour tout  $n \in \mathbb{N}$ ),  $F_n \xrightarrow{d} F$  et  $G_n \xrightarrow{d} G \implies F \prec G$ ;
- 7) (*Invariance à l'ordre des indices*)  $(X_1, X_2) \prec (X'_1, X'_2) \implies (X_2, X_1) \prec (X'_2, X'_1)$ ;
- 8) (*Invariance aux transformations croissantes*) Soit  $a$  une fonction strictement croissante.  $(X_1, X_2) \prec (X'_1, X'_2) \implies (a(X_1), X_2) \prec (a(X'_1), X'_2)$ ;
- 9) (*Invariance aux transformations décroissantes*) Soit  $b$  une fonction strictement décroissante.  $(X_1, X_2) \prec (X'_1, X'_2) \implies (b(X_1), X_2) \prec (b(X'_1), X'_2)$ .

*Remarque 2.7.* Les conditions 2) à 4) impliquent que «  $\prec$  » est une relation d'équivalence.

La première condition, la concordance, se base sur un concept de dépendance important qui n'a pas encore été abordé, soit la dépendance de quadrant.

**Définition 2.8.** Soit  $X = (X_1, X_2)$  un couple de variables aléatoires de fonction de répartition  $F$ . Le vecteur  $X$  (ou la fonction  $F$ ) est dit *de dépendance de quadrant positive* si

$$\mathbb{P}(X_1 > \alpha_1, X_2 > \alpha_2) \geq \mathbb{P}(X_1 > \alpha_1)\mathbb{P}(X_2 > \alpha_2), \text{ pour tout } \alpha_1, \alpha_2 \in \mathbb{R}. \quad (6)$$

*Remarque 2.9.* En réécrivant l'inégalité (6) en utilisant la fonction de survie, on obtient

$$\bar{F}(\alpha_1, \alpha_2) \geq \bar{F}_1(\alpha_1)\bar{F}_2(\alpha_2), \text{ pour tout } \alpha_1, \alpha_2 \in \mathbb{R}. \quad (7)$$

On peut montrer assez facilement que cette dernière est équivalente à

$$F(\alpha_1, \alpha_2) \geq F_1(\alpha_1)F_2(\alpha_2), \text{ pour tout } \alpha_1, \alpha_2 \in \mathbb{R}. \quad (8)$$

La preuve est laissée en exercice.

Ici, nous parlons de dépendance *positive*, car si  $X'_1$  et  $X'_2$  sont des variables aléatoires indépendantes de même distribution marginale que  $X_1$  et  $X_2$ , respectivement, alors  $X_1$  et  $X_2$  sont plus enclines à prendre simultanément de grandes valeurs, ou encore simultanément de petites valeurs, par rapport à  $X'_1$  et  $X'_2$ . De façon analogue, nous pouvons définir la *dépendance de quadrant négative* en inversant le symbole des inégalités (6) à (8).

Du concept de dépendance de quadrant découle un cas particulier d'ordre à dépendance positive bivariable, l'ordre de concordance, que l'on note  $\prec_c$ .

**Définition 2.10.** Soient  $F, G \in \mathcal{C}(F_1, F_2)$  deux fonctions de répartition bivariables. On dit que  $G$  est *plus concordant* que  $F$ , noté  $F \prec_c G$ , si

$$F(x_1, x_2) \leq G(x_1, x_2), \text{ pour tout } x_1, x_2 \in \mathbb{R}.$$

*Remarque 2.11.* De la définition 2.4, celle-ci peut également s'écrire :

$$\bar{F}(x_1, x_2) \leq \bar{G}(x_1, x_2), \text{ pour tout } x_1, x_2 \in \mathbb{R}.$$

Cette définition met en lumière le fait qu'il s'agit de l'ordre à dépendance positive bivariable le plus faible : si  $F \prec G$  pour un ordre à dépendance positive bivariable quelconque, alors  $F \prec_c G$ . On peut alors remarquer que cela constitue la première condition de la définition 2.6.

Le terme *concordant* suggère que si  $(X_1, X_2) \sim F$  et que  $(X'_1, X'_2) \sim G$ , alors  $X'_1$  et  $X'_2$  sont plus sujettes à prendre simultanément de petites valeurs ou de grandes valeurs que  $X_1$  et  $X_2$ . Alors que pour la dépendance de quadrant on comparait un vecteur de variables quelconques à un vecteur de variables indépendantes, la concordance est plutôt une propriété se basant sur la comparaison de vecteurs de variables aléatoires quelconques pour qualifier la dépendance d'une distribution par rapport à une autre.

**Exemple 2.12.** Parmi les distributions  $F \in \mathcal{C}(F_1, F_2)$ ,

- a) la distribution la plus concordante est la *borne supérieure de Fréchet* :

$$F_U(x_1, x_2) = \min\{F_1(x_1), F_2(x_2)\};$$

- b) la distribution la plus discordante est la *borne inférieure de Fréchet* :

$$F_L(x_1, x_2) = \max\{0, F_1(x_1) + F_2(x_2) - 1\}.$$

Il s'agit des bornes de la condition 5) de la définition 2.6.



## 2.2 Mesures d'association

Une des premières mesures de dépendance introduite dans l'étude des statistiques est probablement le coefficient de corrélation de Pearson, une mesure de dépendance linéaire entre deux variables aléatoires quantitatives. Par contre, lorsque nous travaillons avec des variables aléatoires continues non normales, l'utilisation de notions basées sur la linéarité n'est pas toujours appropriée. De plus, contrairement au coefficient de corrélation de Pearson, certaines mesures sont invariantes sous transformations strictement croissantes et valent 1 pour la borne supérieure de Fréchet et -1 pour la borne inférieure de Fréchet, ce qui, en ce sens, les rendent plus intéressantes. C'est le cas entre autres du tau de Kendall  $\tau$  et du rho de Spearman  $\rho_S$ , deux mesures d'association bivariées qui *mesurent* la concordance.

**Définition 2.13.** Soient  $(X_1, X_2)$  et  $(X'_1, X'_2)$  deux paires indépendantes de variables aléatoires de distribution continue  $F \in \mathcal{C}(F_1, F_2)$ . Le *tau de Kendall* se définit comme suit :

$$\tau = \mathbb{P}[(X_1 - X'_1)(X_2 - X'_2) > 0] - \mathbb{P}[(X_1 - X'_1)(X_2 - X'_2) < 0]. \quad (9.a)$$

Autrement dit, il s'agit de la probabilité de concordance entre  $(X_1, X_2)$  et  $(X'_1, X'_2)$  à laquelle nous soustrayons celle de discordance, que l'on peut également réécrire de la manière suivante :

$$\tau = 2\mathbb{P}[(X_1 - X'_1)(X_2 - X'_2) > 0] - 1. \quad (9.b)$$

La proposition qui suit montre comment on peut exprimer le tau de Kendall en fonction de la fonction de répartition conjointe  $F$ .

**Proposition 2.14.** Soient  $(X_1, X_2)$  et  $(X'_1, X'_2)$  deux paires indépendantes de variables aléatoires de distribution continue  $F \in \mathcal{C}(F_1, F_2)$ . Alors, on a

$$\tau = 4 \int F dF - 1. \quad (9.c)$$

*Démonstration.* (Cette démonstration est basée sur la démonstration du théorème 5.1.1 de [Nel06].) On a  $\mathbb{P}[(X_1 - X'_1)(X_2 - X'_2) > 0] = \mathbb{P}(X_1 > X'_1, X_2 > X'_2) + \mathbb{P}(X_1 < X'_1, X_2 < X'_2)$ , des probabilités qui sont évaluées de la manière suivante :

$$\begin{aligned} \mathbb{P}(X_1 < X'_1, X_2 < X'_2) &= \iint_{\mathbb{R}^2} \mathbb{P}(X_1 < x_1, X_2 < x_2) dF(x_1, x_2) \\ &= \iint_{\mathbb{R}^2} F(x_1, x_2) dF(x_1, x_2) \end{aligned}$$

et

$$\begin{aligned}
\mathbb{P}(X_1 > X'_1, X_2 > X'_2) &= \iint_{\mathbb{R}^2} \mathbb{P}(X_1 > x_1, X_2 > x_2) dF(x_1, x_2) \\
&= \iint_{\mathbb{R}^2} \bar{F}(x_1, x_2) dF(x_1, x_2) \\
&= \iint_{\mathbb{R}^2} (1 - F_1(x_1) - F_2(x_2) + F(x_1, x_2)) dF(x_1, x_2) \\
&= 1 - \iint_{\mathbb{R}^2} F_1(x_1) dF(x_1, x_2) - \iint_{\mathbb{R}^2} F_2(x_2) dF(x_1, x_2) \\
&\quad + \iint_{\mathbb{R}^2} F(x_1, x_2) dF(x_1, x_2) \\
&= 1 - \frac{1}{2} - \frac{1}{2} + \iint_{\mathbb{R}^2} F(x_1, x_2) dF(x_1, x_2) \\
&= \iint_{\mathbb{R}^2} F(x_1, x_2) dF(x_1, x_2),
\end{aligned}$$

où l'avant-dernière égalité est obtenue puisque  $F_1(X_1), F_2(X_2) \sim U(0,1)$  et donc l'espérance vaut  $1/2$ . Le résultat suit de l'équation (9.b).  $\square$

Comme le tau de Kendall, le rhô de Spearman est également basé sur la concordance et la discordance.

**Définition 2.15.** Soient  $(X_1, X_2)$ ,  $(X'_1, X'_2)$  et  $(X''_1, X''_2)$  trois paires indépendantes de variables aléatoires de distribution continue  $F \in \mathcal{C}(F_1, F_2)$ . Le *rhô de Spearman* se définit comme suit :

$$\rho_S = 3\mathbb{P}[(X_1 - X'_1)(X_2 - X''_2) > 0] - \mathbb{P}[(X_1 - X'_1)(X_2 - X''_2) < 0]. \quad (10.a)$$

L'équation (10.a) peut facilement se réécrire comme

$$\rho_S = 6\mathbb{P}[(X_1 - X'_1)(X_2 - X''_2) > 0] - 3. \quad (10.b)$$

**Exercice 2.16.** Montrer que le rhô de Spearman peut également s'écrire de la manière suivante :

$$\rho_S = 12 \iint F_1(x_1) F_2(x_2) dF(x_1, x_2) - 3. \quad (10.c)$$

Une fois les copules introduites, nous verrons comment écrire ces mesures uniquement à l'aide de celles-ci. Bien que les valeurs du tau de Kendall et du rhô de Spearman se situent entre -1 et 1, la connaissance de l'une permet d'avoir une bonne idée sur celle de l'autre grâce aux inégalités du théorème suivant.

**Théorème 2.17.** (*Théorèmes 5.1.10 et 5.1.11 de [Nel06]*) On a

$$|3\tau - 2\rho| \leq 1$$

ainsi que

$$\frac{1 + \rho_S}{2} \geq \left(\frac{1 + \tau}{2}\right)^2 \quad \text{et} \quad \frac{1 - \rho_S}{2} \geq \left(\frac{1 - \tau}{2}\right)^2.$$

En isolant  $\rho_S$  dans chacune des inégalités du théorème 2.17 et en combinant les résultats, on obtient le corollaire qui suit.

**Corollaire 2.18.** *Selon le signe de  $\tau$ , nous avons*

$$\begin{cases} \frac{3\tau - 1}{2} \leq \rho_S \leq \frac{1 + 2\tau - \tau^2}{2} & \text{si } \tau \geq 0 \\ \frac{\tau^2 + 2\tau - 1}{2} \leq \rho_S \leq \frac{1 + 3\tau}{2} & \text{si } \tau \leq 0. \end{cases}$$

Ainsi, pour toute paire de variables aléatoires continues  $(X_1, X_2)$ , les valeurs de  $\tau$  et  $\rho_S$  doivent se situer dans la région ombragée de la figure 2.

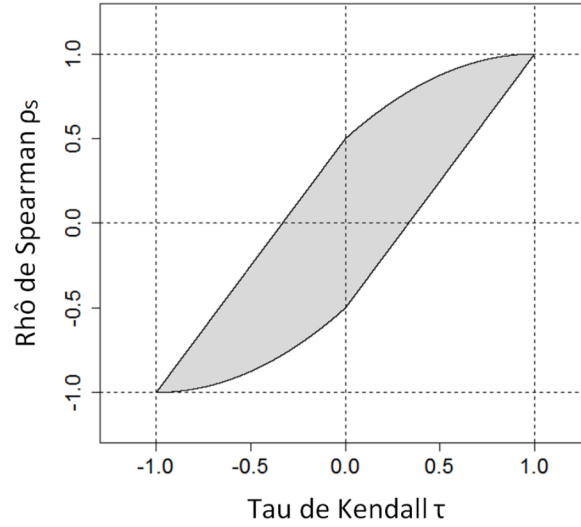


FIGURE 2 – Région contenant les valeurs de  $\tau$  et de  $\rho_S$ , l'une en fonction de l'autre

Jusqu'à présent, les concepts de dépendance et les mesures présentés servaient à décrire comment les valeurs extrêmes d'une variable aléatoire se comportent avec celles d'une autre. Dans un autre ordre d'idées, la dépendance de queue se rapporte plutôt à la dépendance des quadrants inférieur gauche et supérieur droit d'une distribution bivariée, ce qui permet de décrire la dépendance de valeurs extrêmes (qui se rapportent principalement aux queues d'une distribution). Le paramètre utilisé pour une telle mesure sera  $\lambda$ . Pour ce faire, nous avons besoin des inverses des fonctions marginales. Par contre, si ces dernières ne sont pas strictement croissantes, nous devons utiliser le *quasi-inverse*, car la fonction inverse ne sera pas correctement définie.

**Définition 2.19.** Soit  $F$  une fonction de répartition. Le *quasi-inverse*  $F^{(-1)} : \mathbb{I} \rightarrow \mathbb{R}$  de  $F$  est une fonction définie comme suit :

$$F^{(-1)}(t) = \inf\{x \mid F(x) \geq t\}. \quad (11)$$

*Remarque 2.20.* Si  $F$  est strictement croissante, alors le quasi-inverse correspond à l'inverse usuel de  $F$ , que nous noterons  $F^{-1}$ .

**Définition 2.21.** Soient  $X_1$  et  $X_2$  deux variables aléatoires de fonction de répartition univariée respective  $F_1$  et  $F_2$ . Si les limites existent, un paramètre permettant de mesurer la dépendance de queue à droite est

$$\lambda_U = \lim_{t \rightarrow 1^-} \mathbb{P}[X_2 > F_2^{(-1)}(t) | X_1 > F_1^{(-1)}(t)] \quad (12)$$

et un permettant de mesurer la dépendance de queue à gauche est

$$\lambda_L = \lim_{t \rightarrow 0^+} \mathbb{P}[X_2 \leq F_2^{(-1)}(t) | X_1 \leq F_1^{(-1)}(t)]. \quad (13)$$

On verra dans la prochaine section qu'il est également possible d'écrire ces mesures en termes de la copule sous-jacente, là où l'interprétation sera plus facile.

**Exercice 2.22.** Soient  $F, G \in \mathcal{C}(F_1, F_2)$  deux fonctions de répartition bivariée. Montrer que  $F \prec_c G$  implique que (i)  $\tau(F) \leq \tau(G)$ , (ii)  $\rho_S(F) \leq \rho_S(G)$  et (iii)  $\lambda(F) \leq \lambda(G)$ .

### 3 Copules

Nous voici rendus au sujet principal de cet article : les copules. Celles-ci permettent de résumer la structure de dépendance d'une distribution conjointe en la séparant des comportements marginaux. Dans cette section, nous formalisons ce concept en donnant une définition précise, en énonçant le fameux théorème de Sklar ainsi qu'en présentant quelques propriétés et résultats.

#### 3.1 Éléments de base

*Notation 3.1.* Posons  $I = [0,1] \subset \mathbb{R}$ . Ainsi,  $I^2$  fera référence au carré unitaire  $[0,1] \times [0,1] \subset \mathbb{R}^2$ .

Puisque la distribution uniforme est intimement liée aux copules, rappelons la fonction de répartition univariée d'une variable aléatoire de loi uniforme sur l'intervalle  $I$ . Soit  $U \sim U(0,1)$ . Alors, on a

$$\mathbb{P}(U \leq u) = \begin{cases} 0 & \text{si } u \leq 0 \\ u & \text{si } 0 \leq u \leq 1 \\ 1 & \text{si } u \geq 1. \end{cases} \quad (14)$$

**Définition 3.2.** Une *copule bivariée* est une fonction  $C : I^2 \rightarrow I$  qui est définie par les conditions suivantes :

- 1)  $C$  est attachée<sup>1</sup>, c'est-à-dire

$$C(u,0) = 0 = C(0,v), \text{ pour tout } u,v \in I; \quad (15.a)$$

---

<sup>1</sup>De l'anglais « grounded ».

2) Les marges sont uniformes, c'est-à-dire

$$C(u,1) = u \text{ et } C(1,v) = v, \text{ pour tout } u,v \in I; \quad (15.b)$$

3)  $C$  est une fonction 2-croissante sur  $I^2$ , c'est-à-dire pour tout  $u_1, u_2, v_1, v_2 \in I$  tels que  $u_1 \leq u_2$  et  $v_1 \leq v_2$ , on a

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0. \quad (15.c)$$

En ramenant la normalisation de distributions que Hoeffding avait faite sur la région  $I^2$  plutôt que sur  $[-1/2, 1/2]^2$ , cela l'aurait conduit à la découverte des copules [Nel06].

En comparant la définition de copules avec celle d'une fonction de répartition bivariable, nous constatons qu'il s'agit d'une restriction de cette dernière en limitant le domaine à  $I^2$  et en exigeant que les marges soient uniformes. Ainsi, de nombreux parallèles peuvent être effectués avec plusieurs notions de dépendance présentées dans la section précédente.

Par exemple, il est possible de déterminer les copules extrêmes, soient les bornes de Fréchet. Posons  $C^-(u,v) = \max(0, u + v - 1)$  et  $C^+(u,v) = \min(u,v)$  (à comparer avec l'exemple 2.12).

**Théorème 3.3.** *Soit  $C$  une copule. Alors, pour tout  $(u,v) \in I^2$ , on a*

$$C^-(u,v) \leq C(u,v) \leq C^+(u,v). \quad (16)$$

*Démonstration.* Soit  $(u,v) \in I^2$ . Puisque  $C(u,v) \leq C(u,1) = u$  et  $C(u,v) \leq C(1,v) = v$ , alors  $C(u,v) \leq \min(u,v) = C^+(u,v)$ , d'où la deuxième inégalité. Pour démontrer la première inégalité, nous utilisons le fait que  $C(u,v) \geq 0$  et l'inégalité (15.c) :

$$\begin{aligned} C(1,1) - C(1,v) - C(u,1) + C(u,v) &\geq 0 \implies 1 - v - u + C(u,v) \geq 0 \\ &\implies C(u,v) \geq u + v - 1 \\ &\implies C(u,v) \geq \max(0, u + v - 1) \\ &\implies C(u,v) \geq C^-(u,v), \end{aligned}$$

d'où le résultat. □

La figure 3 illustre les copules extrêmes. Ainsi, de la définition et du fait qu'une copule est uniformément continue sur son domaine (théorème 2.2.4. de [Nel06]), le graphique de toute copule  $C$  est un quadrilatère de surface continue compris dans le cube unitaire  $I^3$ , limité par les copules extrêmes et ayant pour sommets les points  $(0,0,0)$ ,  $(1,0,0)$ ,  $(0,1,0)$  et  $(1,1,1)$ .

Une autre manière de représenter graphiquement une copule est en utilisant les courbes de niveau, où une courbe est déterminée par l'ensemble des points  $(u,v) \in I^2$  tels que  $C(u,v) = t$  pour un certain  $t \in I$ . En se basant sur celles des bornes de Fréchet présentées à la figure 4, nous pouvons déduire que, pour toute copule  $C$ , sa courbe de niveau pour un certain  $t_0 \in I$  doit se trouver dans la région ombragée de la figure 5.

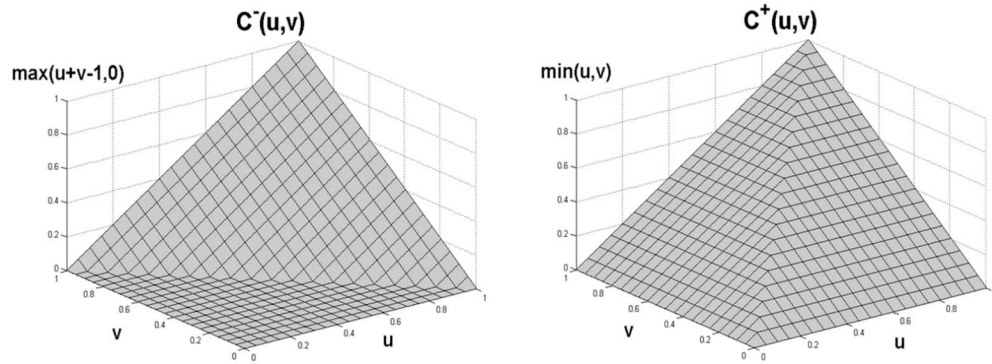


FIGURE 3 – Graphique des copules extrêmes

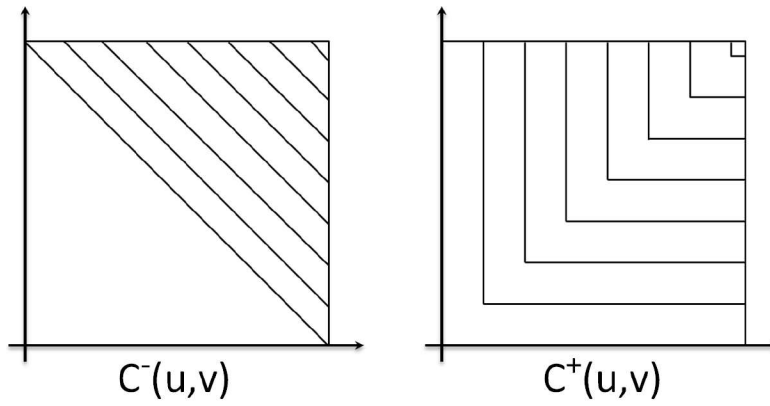


FIGURE 4 – Courbes de niveau des copules extrêmes

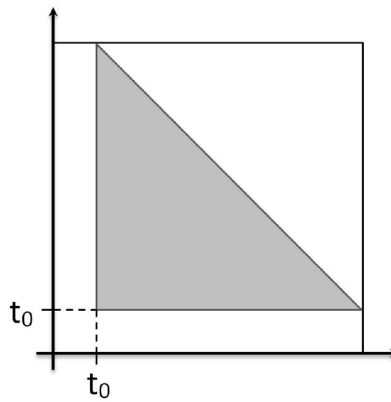


FIGURE 5 – Région contenant la courbe de niveau  $C(u,v) = t_0$  d'une certaine copule  $C$

### 3.2 Théorème de Sklar

De la définition 3.2, il est possible de construire une copule en s'assurant que les trois conditions de la définition soient respectées. Une autre manière consiste à utiliser la fonction de répartition bidimensionnelle. Le théorème de Sklar met en lumière comment il est possible de relier le principe de la copule à la fonction de répartition bivariée et aux lois marginales.

**Théorème 3.4.** (*Théorème 2.3.3 de [Nel06]*) Soit  $F$  une fonction de répartition conjointe de marges  $F_1$  et  $F_2$ . Alors, il existe une copule  $C : I^2 \rightarrow I$  telle que pour tout  $(x, y) \in \mathbb{R}^2$ , on a :

$$F(x, y) = C(F_1(x), F_2(y)). \quad (17)$$

De plus, si  $F_1$  et  $F_2$  sont continues, alors  $C$  est unique.  $\square$

L'équation (17) permet d'exprimer une fonction de répartition bivariée  $F$  à l'aide d'une copule et de lois marginales  $F_1$  et  $F_2$ . À l'inverse, il est également possible d'exprimer une copule à partir de  $F$  et des inverses (ou quasi-inverses) des fonctions de répartition marginales.

**Corollaire 3.5.** Soit  $F$  une fonction de répartition conjointe de marges continues  $F_1$  et  $F_2$  et  $C$  la copule associée à  $F$ . Alors, pour  $(u, v) \in I^2$ , on a

$$\begin{aligned} C(u, v) &= C(F_1(x_1), F_2(x_2)) \\ &= F(F_1^{(-1)}(u), F_2^{(-1)}(v)). \end{aligned} \quad (18)$$

Ainsi, du corollaire 3.5, nous avons une formule de départ pour construire une copule bivariée à partir d'une fonction de distribution conjointe et de lois marginales.

**Exemple 3.6.** Soit la distribution logistique bivariée de Gumbel [Gum61] dont la fonction de répartition conjointe est donnée par

$$F(x_1, x_2) = (1 + e^{-x_1} + e^{-x_2})^{-1}, \text{ pour tout } (x_1, x_2) \in \mathbb{R}^2.$$

Des équations (3.a) et (3.b), nous obtenons les fonctions de distribution marginales  $F_1(x_1) = (1 + e^{-x_1})^{-1}$  et  $F_2(x_2) = (1 + e^{-x_2})^{-1}$ . Comme les fonctions inverses sont respectivement  $F_1^{-1}(u) = -\ln\left(\frac{1}{u} - 1\right)$  et  $F_2^{-1}(v) = -\ln\left(\frac{1}{v} - 1\right)$ , en se basant sur l'équation (18), la copule associée à  $F(x_1, x_2)$  s'écrit

$$\begin{aligned} C(u, v) &= F(F_1^{(-1)}(u), F_2^{(-1)}(v)) \\ &= \left[ 1 + e^{\ln\left(\frac{1}{u} - 1\right)} + e^{\ln\left(\frac{1}{v} - 1\right)} \right]^{-1} \\ &= \frac{uv}{u + v - uv}. \end{aligned}$$

### 3.3 Propriétés

Cette section met en lumière quelques propriétés importantes de la théorie des copules. Un des aspects qui font des copules un outil statistique avantageux relève de leur caractère prévisible lorsque sont appliquées aux variables aléatoires des transformations strictement monotones.

*Notation 3.7.* Pour préciser que la copule  $C$  est associée aux variables  $X_1$  et  $X_2$ , nous la noterons  $C_{X_1, X_2}$ .

**Théorème 3.8.** *Soient  $X_1$  et  $X_2$  deux variables aléatoires continues ayant pour copule  $C_{X_1, X_2}$ . Si  $\alpha$  et  $\beta$  sont deux fonctions strictement croissantes, alors  $C_{X_1, X_2} = C_{\alpha(X_1), \beta(X_2)}$ .*

*Démonstration.* Soient  $\alpha$  et  $\beta$  deux fonctions strictement croissantes et  $X_1$  et  $X_2$  deux variables aléatoires continues ayant pour copule  $C_{X_1, X_2}(u, v) = F(x_1, x_2)$  et pour fonctions de répartition marginales respectivement  $F_1$  et  $F_2$ . Notons les fonctions de distribution univariées de  $\alpha(X_1)$  et  $\beta(X_2)$  respectivement  $G_1$  et  $G_2$  et la loi conjointe  $G$ . Puisque  $\alpha$  et  $\beta$  sont strictement croissantes et que  $X_1$  et  $X_2$  sont des variables aléatoires continues, pour tout  $(u, v) \in \mathbb{I}^2$ , on a

$$\begin{aligned}
 C_{\alpha(X_1), \beta(X_2)}(u, v) &= C_{\alpha(X_1), \beta(X_2)}(G_1(x_1), G_2(x_2)) \\
 &= G(x_1, x_2) \\
 &= \mathbb{P}(\alpha(X_1) \leq x_1, \beta(X_2) \leq x_2) \\
 &= \mathbb{P}(X_1 \leq \alpha^{-1}(x_1), X_2 \leq \beta^{-1}(x_2)) \\
 &= F(\alpha^{-1}(x_1), \beta^{-1}(x_2)) \\
 &= C_{X_1, X_2}(F_1(\alpha^{-1}(x_1)), F_2(\beta^{-1}(x_2))) \\
 &= C_{X_1, X_2}(G_1(x_1), G_2(x_2)). \quad \square
 \end{aligned}$$

Dans la littérature des copules, ce théorème est connu comme étant le théorème d'invariance sous transformations strictement croissantes. Bien que l'invariance soit assurée seulement pour une croissance stricte des deux fonctions appliquées, il existe un résultat dans le cas où  $\alpha$  ou  $\beta$  sont plutôt décroissantes, le comportement de la copule étant ainsi prévisible (voir théorème 2.4.4 de [Nel06]). Comme le tau de Kendall  $\tau$  et le rhô de Spearman  $\rho_S$  sont invariants aux transformations strictement croissantes (la preuve est laissée en exercice), à la définition présentée dans la section 2.2 s'ajoute celle en termes de copules, où  $C$  est la copule associée  $F$  :

$$\tau_C = 4 \iint_{\mathbb{I}^2} C(u, v) dC(u, v) - 1 \quad (19.a)$$

$$= 4E(C(U, V)) - 1 \quad (19.b)$$

et

$$\rho_{SC} = 12 \iint_{\mathbb{I}^2} uv dC(u, v) - 3. \quad (20)$$



Un autre concept de dépendance qui a été abordé précédemment est celui de la dépendance de queue, qui, rappelons-le, mesure la probabilité que des valeurs extrêmes se réalisent simultanément. Tout comme  $\tau$  et  $\rho_S$ ,  $\lambda_L$  et  $\lambda_U$  peuvent s'écrire en termes de copules. Comme la définition de ces deux derniers paramètres se base sur des probabilités conditionnelles, le lemme suivant les mettra en relation avec les copules, ce qui facilitera la démonstration du théorème 3.10.

**Lemme 3.9.** *Soient  $U$  et  $V$  deux variables aléatoires uniformes sur  $I$ , et  $C$  la copule associée. Pour  $u, v \in I$ , on a*

$$\mathbb{P}(U \leq u \mid V \leq v) = \frac{C(u, v)}{v} \text{ et } \mathbb{P}(U > u \mid V > v) = \frac{1 - u - v + C(u, v)}{1 - v}.$$

*Démonstration.* On a

$$\begin{aligned} \mathbb{P}(U \leq u \mid V \leq v) &= \frac{\mathbb{P}(U \leq u, V \leq v)}{\mathbb{P}(V \leq v)} = \frac{C(u, v)}{v} \\ \text{et } \mathbb{P}(U > u \mid V > v) &= \frac{\mathbb{P}(U > u, V > v)}{\mathbb{P}(V > v)} \\ &= \frac{1 - \mathbb{P}(U \leq u) - \mathbb{P}(V \leq v) + \mathbb{P}(U \leq u, V \leq v)}{1 - \mathbb{P}(V \leq v)} \\ &= \frac{1 - u - v + C(u, v)}{1 - v}. \end{aligned} \quad \square$$

**Théorème 3.10.** *Soient  $X$  et  $Y$  deux variables aléatoires définies comme dans la définition 2.21. Si les limites des équations (12) et (13) existent, alors*

$$\lambda_U = 2 - \lim_{u \rightarrow 1^-} \frac{1 - C(u, u)}{1 - u} \text{ et } \lambda_L = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u}.$$

*Démonstration.* On a

$$\begin{aligned} \lambda_U &= \lim_{u \rightarrow 1^-} \mathbb{P}[X_2 > F_2^{(-1)}(u) \mid X_1 > F_1^{(-1)}(u)] \\ &= \lim_{u \rightarrow 1^-} \mathbb{P}[F_2(X_2) > u \mid F_1(X_1) > u] \\ &= \lim_{u \rightarrow 1^-} \frac{1 - u - u + C(u, u)}{1 - u} \\ &= 2 - \lim_{u \rightarrow 1^-} \frac{1 - C(u, u)}{1 - u} \\ \text{et } \lambda_L &= \lim_{u \rightarrow 0^+} \mathbb{P}[X_2 \leq F_2^{(-1)}(u) \mid X_1 \leq F_1^{(-1)}(u)] \\ &= \lim_{u \rightarrow 0^+} \mathbb{P}[F_2(X_2) \leq u \mid F_1(X_1) \leq u] \\ &= \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u}. \end{aligned} \quad \square$$

La valeur de ces paramètres de dépendance de queue se situe toujours dans l'intervalle  $I$ . Selon celle-ci, on dira qu'une copule  $C$  possède ou non une dépendance de queue à gauche ou à droite.

**Définition 3.11.** Soit une copule  $C$ .

- a) Si  $\lambda_U$  existe et que sa valeur est comprise dans l'intervalle  $I^*$ , alors  $C$  a une *dépendance de queue à droite*.
- b) Si  $\lambda_L$  existe et que sa valeur est comprise dans l'intervalle  $I^*$ , alors  $C$  a une *dépendance de queue à gauche*.

Autrement dit, si  $\lambda_U > 0$  (respectivement  $\lambda_L > 0$ ), alors il existe une probabilité non nulle qu'une des deux variables aléatoires prenne une valeur supérieure (respectivement inférieure) à un certain  $u$  lorsque l'autre est supérieure (respectivement inférieure) à  $u$ , pour un certain  $u$  arbitrairement près de 1 (respectivement 0) [Joe01].

**Exemple 3.12.** Considérons la copule logistique bivariée de Gumbel telle que définie à l'exemple 3.6. Comme on a

$$\begin{aligned} \lambda_U &= 2 - \lim_{u \rightarrow 1^-} \frac{1 - C(u, u)}{1 - u} & \text{et} & & \lambda_L &= \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u} \\ &= 2 - \lim_{u \rightarrow 1^-} \frac{2}{2 - u} & & & &= \lim_{u \rightarrow 0^+} \left( \frac{u}{2 - u} \right) \left( \frac{1}{u} \right) \\ &= 0 & & & &= \frac{1}{2}, \end{aligned}$$

cette copule possède une dépendance de queue à gauche, mais ne possède pas de dépendance de queue à droite.

Par ailleurs, il n'est pas rare de constater une absence de dépendance entre deux caractères, ce que nous qualifions d'indépendance. Si  $X$  et  $Y$  sont deux variables aléatoires indépendantes, nous savons que leur fonction de répartition conjointe s'exprimera comme le produit de leur fonction de répartition marginale respective. Ainsi, le prochain théorème montre la copule qui sera associée à l'indépendance, que nous noterons  $C^\perp(u, v)$  : tout couple de variables aléatoires indépendantes satisfera à l'équation qui y sera énoncée.

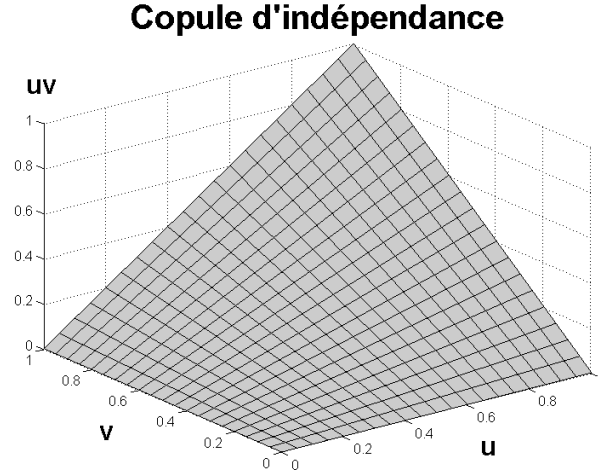
**Théorème 3.13.** Soient  $X_1$  et  $X_2$  deux variables aléatoires continues.  $X_1$  et  $X_2$  sont indépendantes si et seulement si  $C_{X_1, X_2}(u, v) = C^\perp(u, v) = uv$ .

*Démonstration.* La preuve, laissée en exercice, découle du fait que  $F(x, y) = F_1(x)F_2(y)$  et des relations présentées aux équations (3.a), (3.b) et (17).  $\square$

Comme la copule d'indépendance, illustrée à la figure 6, s'exprime par le produit  $uv$ , nous l'appelons parfois *copule produit* et la notons  $\pi(u, v)$ .

## 4 Exemples de copules

Une fois les données recueillies, pour déterminer les relations entre les variables, il faut choisir la copule qui résume le mieux la structure de dépendance à l'aide

FIGURE 6 – Copule d'indépendance  $C^{\perp}(u, v)$ 

de tests d'adéquation pour le type de copules ainsi que le ou les paramètres en cause. Nous voyons ici quelques exemples de copules couramment utilisées en présentant leurs caractéristiques. Nous illustrons également leur densité et simulons un exemple de nuage de points lui étant associé. La densité d'une copule, notée  $c(u, v)$ , se définit à l'aide de la fonction copule  $C(u, v)$  comme suit :

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v). \quad (22)$$

À partir de celle-ci, il est possible d'obtenir la fonction de densité [Cla10] :

$$f(x_1, x_2) = c(F_1(x_1), F_2(x_2)) \times f_1(x_1) \times f_2(x_2), \quad (23)$$

où  $f_1(\cdot)$  et  $f_2(\cdot)$  sont les fonctions de densité marginales. De l'équation (22), on peut facilement calculer que la densité de la copule d'indépendance  $C^{\perp}(u, v)$  vaut 1 sur tout  $\mathbb{I}^2$ .

#### 4.1 Copules elliptiques

Du théorème de Sklar, il est possible d'extraire d'une distribution elliptique une copule dite elliptique. Pour un couple de variables aléatoires  $X = (X_1, X_2)$ , on dira que  $X$  suit une loi elliptique bivariable si et seulement si on peut écrire  $X = \mu + R\Sigma^{1/2}\mathcal{U}$ , où  $\mu$  est la moyenne, si elle existe,  $R$  une variable aléatoire positive indépendante de  $\mathcal{U}$ ,  $\Sigma$  une matrice de forme (matrice de variance-covariance, si elle existe) et  $\mathcal{U}$  une variable aléatoire uniformément distribuée sur le disque unitaire de  $\mathbb{R}^2$  (voir [Mui82] pour plus de détails). Les courbes de niveau des densités de telles distributions forment des ellipses d'excentricité constante [Nel06]. Dans cette famille se trouvent entre autres les copules gaussiennes et les copules Student.

#### 4.1.1 Copule gaussienne

La copule gaussienne fait partie de la famille elliptique des copules bivariées à un paramètre. Utiliser cette copule est conséquent avec la mesure de dépendance obtenue par le coefficient de corrélation linéaire  $\rho$  [Cla10].

**Définition 4.1.** Soient  $\rho$  le coefficient de corrélation,  $\Phi$  la distribution normale  $\mathcal{N}(0,1)$  et  $\Phi_\rho$  la distribution normale bivariée de matrice de corrélation associée à  $\rho$ . Alors, la *copule gaussienne bivariée*  $C_G$  est définie de la manière suivante :

$$C_G(u,v;\rho) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)). \quad (24)$$

La figure 7 montre le graphique de la densité de trois copules gaussiennes qui diffèrent par leur valeur de  $\rho$ .

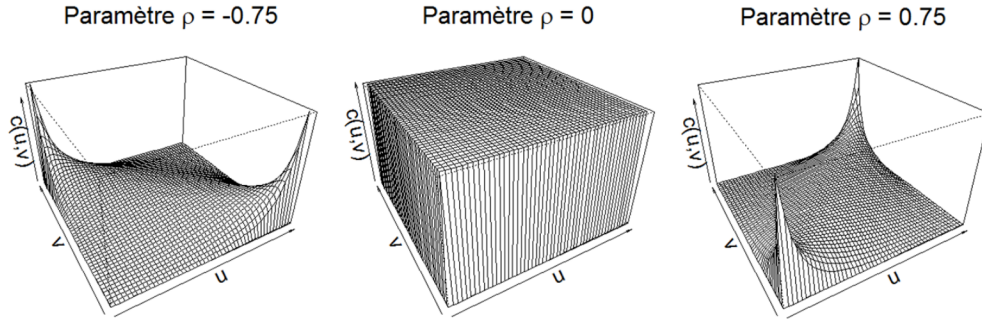


FIGURE 7 – Densité  $c_G(u,v;\rho)$  de 3 copules gaussiennes selon  $\rho$

On peut remarquer que, pour  $\rho = 0$ , on a  $C_G = C^\perp$ . Pour ces mêmes valeurs de  $\rho$ , la figure 8, quant à elle, illustre ce à quoi pourrait ressembler un nuage de points d'une copule gaussienne.

Ne présentant pas de dépendance de queue (sauf pour  $\rho = 1$ ), la copule gaussienne n'est pas adaptée pour modéliser l'interaction de valeurs extrêmes. Dans ce cas, mieux vaudra utiliser d'autres copules, tout comme pour caractériser une dépendance non linéaire.

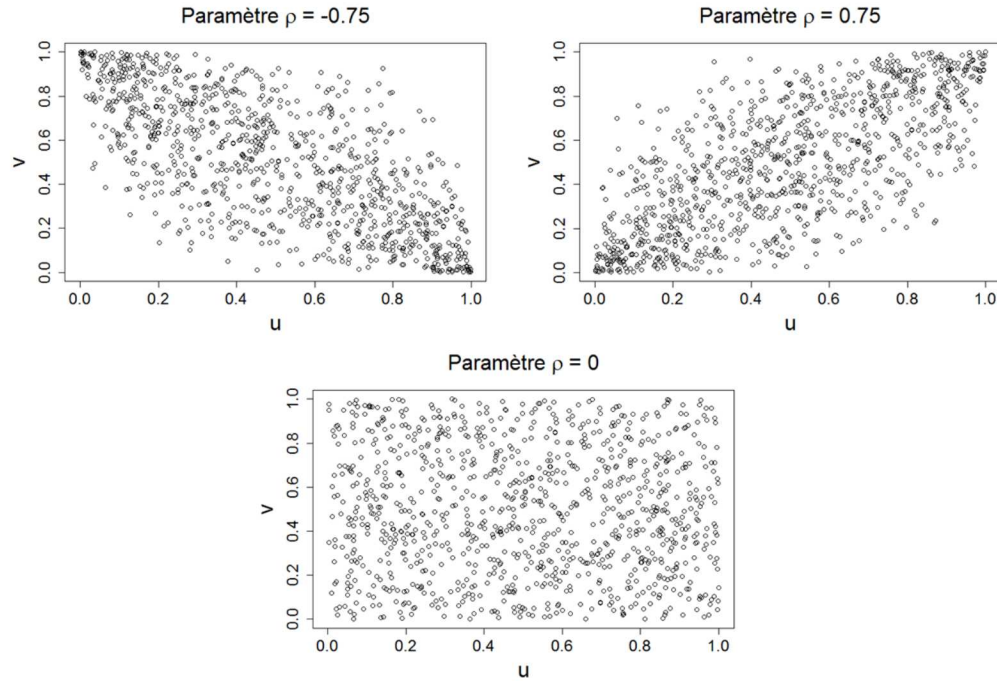
**Exercice 4.2.** Montrer que dans le cas de copules gaussiennes, le tau de Kendall s'écrit de la manière suivante :  $\tau = \frac{2}{\pi} \arcsin(\rho)$ .

#### 4.1.2 Copule Student

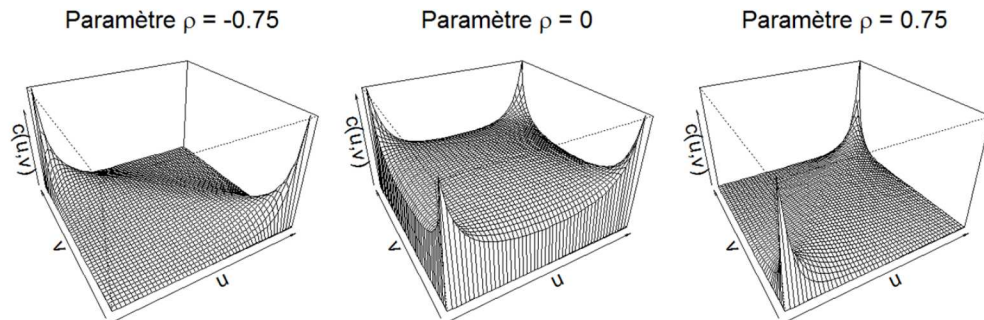
Construite de manière similaire à la copule gaussienne, la copule Student est extraite à partir de la distribution de Student bivariée.

**Définition 4.3.** Soient  $\rho$  le coefficient de corrélation,  $T$  la distribution de Student bivariée standardisée et  $\Phi_{\rho,\kappa}$  la distribution de Student bivariée de matrice de corrélation associée à  $\rho$  et de degré de liberté  $\kappa \geq 3$ . Alors, la *copule Student bivariée*  $C_T$  est définie de la manière suivante :

$$C_T(u,v;\rho,\kappa) = T_{\rho,\kappa}(T^{-1}(u), T^{-1}(v)). \quad (25)$$

FIGURE 8 – Nuages de points de trois copules gaussiennes selon  $\rho$ 

En comparaison avec la copule gaussienne, la figure 9 montre le graphique de la densité de trois copules Student de degré de liberté  $\kappa = 5$  qui diffèrent par leur valeur de  $\rho$ .

FIGURE 9 – Densité  $c_T(u, v; \rho, \kappa = 5)$  de 3 copules Student selon  $\rho$ 

Contrairement à la copule gaussienne, la copule Student montre des dépendances de queue à droite et à gauche pour  $\rho \neq -1$  [Cla10]. De plus, on remarque que l'indépendance n'est pas caractérisée par une valeur nulle de  $\rho$ , car la densité  $c_T(u, v)$  ne vaut pas uniquement 1 sur  $I^2$ . Ce sera le cas notamment pour de petites valeurs de degré de liberté  $\kappa$ . Ceci est également visible en observant les nuages de points de la figure 10, où, pour  $\rho = 0$ , on remarque une plus grande

concentration de points dans les coins du graphique.

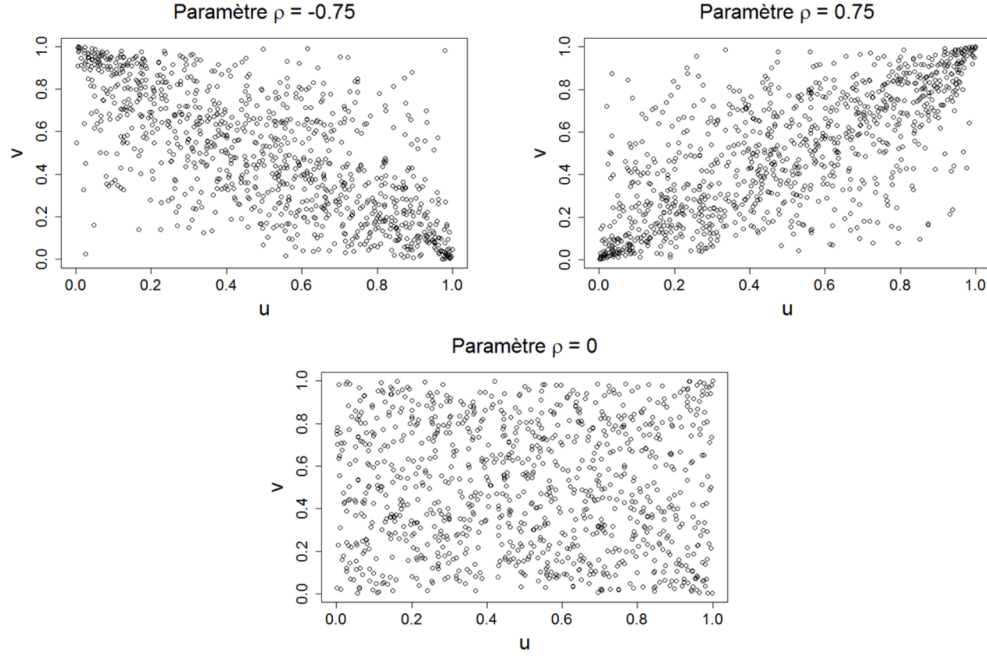


FIGURE 10 – Nuages de points de trois copules Student selon  $\rho$

## 4.2 Copules archimédiennes

Les copules archimédiennes constituent une classe très importante de copules, principalement en raison de la facilité avec laquelle on peut les construire, du nombre important de familles qui appartiennent à celles-ci et des propriétés intéressantes qu'elles possèdent [Nel06].

**Définition 4.4.** Soit  $\varphi : [0,1] \rightarrow [0,\infty)$  une fonction telle que, pour tout  $u \in [0,1]$ ,  $\varphi(1) = 0$ ,  $\varphi'(u) < 0$  et  $\varphi''(u) > 0$ . Les *copules archimédiennes* sont définies comme suit :

$$C(u,v) = \begin{cases} \varphi^{-1}(\varphi(u) + \varphi(v)) & \text{si } \varphi(u) + \varphi(v) \leq \varphi(0) \\ 0 & \text{sinon.} \end{cases} \quad (26)$$

La fonction  $\varphi$  sera dite *fonction génératrice* de la copule.

Dans plusieurs cas, la fonction génératrice facilite l'étude d'une copule, par exemple en ce qui concerne l'ordre stochastique (voir section 4.4 de [Nel06] pour plus de détails).

**Théorème 4.5. (Corollaire 5.1.4 de [Nel06])** Soient  $X_1$  et  $X_2$  deux variables aléatoires de copule archimédienne  $C$  et de fonction génératrice  $\varphi$ . Alors, le tau de Kendall s'écrit

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt.$$

### 4.2.1 Copule Clayton

Parmi cette classe de copule, on retrouve notamment la copule Clayton. De générateur  $\varphi_\theta(t) = \frac{1}{\theta}(t^{-\theta} - 1)$ , pour  $\theta \in [-1, \infty) \setminus \{0\}$ , cette copule s'écrit

$$C_\theta(u, v) = [\max(u^{-\theta} + v^{-\theta} - 1, 0)]^{-1/\theta}.$$

La figure 11 illustre la densité de la copule Clayton ainsi qu'un nuage de points associé à celle-ci.

On remarque que la copule Clayton est asymétrique et ne détecte pas de dépendance pour de grandes valeurs, c'est-à-dire  $\lambda_U = 0$  et  $\lambda_L \neq 0$ . Le tau de Kendall associé à une copule Clayton de fonction génératrice  $\varphi_\theta(t)$  est  $\tau = \frac{\theta}{\theta + 2}$  (la preuve est laissée en exercice).

### 4.2.2 Copule Frank

La copule Frank, quant à elle, a pour générateur  $\varphi_\theta(t) = -\ln\left(\frac{e^{\theta t} - 1}{e^\theta - 1}\right)$ , pour  $\theta \in \mathbb{R}^*$ , et comme formule

$$C_\theta(u, v) = -\frac{1}{\theta} \ln \left( 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right).$$

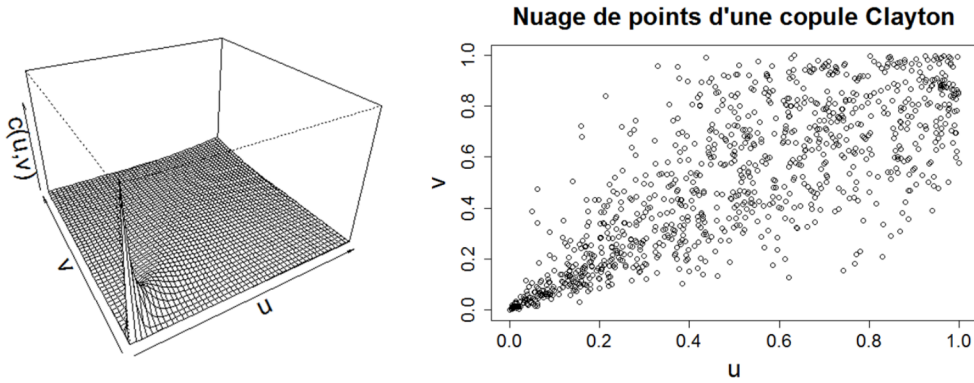
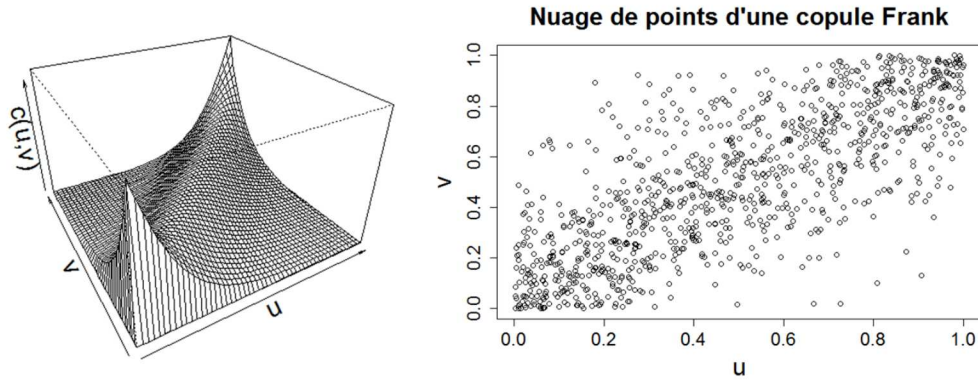


FIGURE 11 – Densité et nuage de points d'une copule Clayton pour  $\theta = 3$

De la même manière que pour la copule Clayton, la figure 12 illustre graphiquement la copule Frank.

Du nuage de points, on remarque que cette copule symétrique présente une plus forte concentration de points aux coins  $(0,0)$  et  $(1,1)$ , comparativement au reste du nuage où l'étalement des points est plus prononcé, ce que corrobore le graphique de la densité.



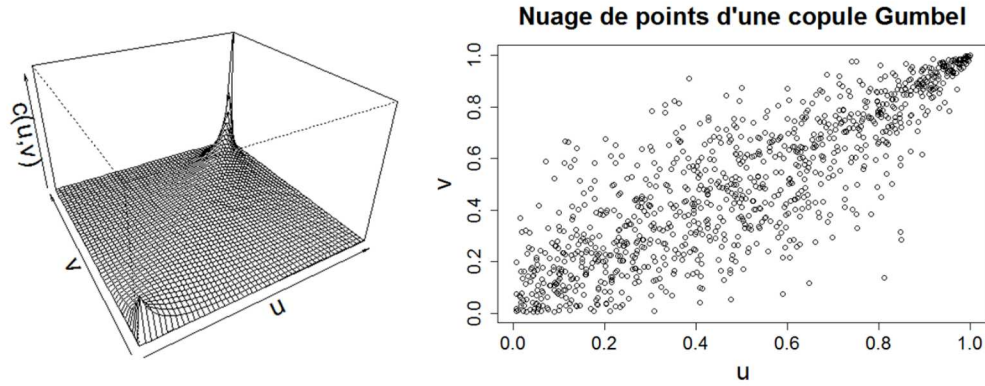
FIGURE 12 – Densité et nuage de points d'une copule Frank pour  $\theta = 7$ 

#### 4.2.3 Copule Gumbel

La dernière copule archimédienne présentée ici est la copule Gumbel, parfois appelée Gumbel-Hougaard, dont la fonction génératrice est  $\varphi_\theta(t) = (-\ln t)^\theta$  pour  $\theta \geq 1$ . Cette fonction copule est définie par l'expression suivante :

$$C_\theta(u, v) = e^{-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}}.$$

La figure 13 est l'équivalent des figures 11 et 12 pour la copule de Gumbel.

FIGURE 13 – Densité et nuage de points d'une copule Gumbel pour  $\theta = 3$ 

Étant utile pour déceler des dépendances de queue à gauche et à droite, cette copule aura donc des valeurs de  $\lambda_U$  et  $\lambda_L$  non nulles. Pour cette copule, on a  $\tau = 1 - \frac{1}{\theta}$  (la preuve est laissée en exercice).

**Exercice 4.6.** La copule produit est un cas particulier d'une copule archimédienne. Trouvez la fonction génératrice  $\varphi_\theta(t)$  qui lui est associée.



### 4.3 Copules de valeurs extrêmes

Comme leur nom l'indique, les copules de valeurs extrêmes sont issues de la théorie de valeurs extrêmes. Dans cette dernière, on considère les variables aléatoires de maximum suivantes :

$$M_{(n,X)} = \max_{1 \leq i \leq n} (X_i) \text{ et } M_{(n,Y)} = \max_{1 \leq j \leq n} (Y_j).$$

On étudie alors la distribution de probabilité de la variable

$$M_n = (M_{(n,X)}, M_{(n,Y)})$$

et les cas pour lesquels cette variable, qu'on a renormalisée, ne suit pas une loi dégénérée. Pour plus de détails, se référer à [Joe01].

**Définition 4.7.** Soit  $k$  une constante réelle positive. Une *copule de valeurs extrêmes* est une copule qui vérifie la relation suivante :

$$C(u^k, v^k) = C^k(u, v). \quad (27)$$

**Exemple 4.8.** La copule de Gumbel est une copule de valeurs extrêmes. En effet, pour une constante  $k$  réelle positive, on a

$$\begin{aligned} C(u^k, v^k) &= e^{-[(-\ln(u^k))^\theta + (-\ln(v^k))^\theta]^{1/\theta}} \\ &= e^{-[k^\theta [(-\ln u)^\theta + (-\ln v)^\theta]]^{1/\theta}} \\ &= e^{-k [(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}} \\ &= \left( e^{-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}} \right)^k \\ &= C^k(u, v). \end{aligned}$$

## 5 Conclusion

Bien que le domaine des copules soit récent, plusieurs applications y sont rattachées, notamment en finance où l'utilisation des copules est parfois bien plus appropriée que certaines méthodes statistiques traditionnelles de corrélation (voir [Bou10]).

## Références

- [Bou10] Pierre BOUVIER : Application des copules à la finance des marchés. Thèse de doctorat, Université du Québec à Montréal, Montréal, Février 2010.
- [Cla10] Pierre CLAUS : Théorie des copules. Notes de cours, École nationale de la statistique et de l'analyse de l'information (Ensa), Filière Gestion des risques et ingénierie financière, 2007-2010.

- [Fré51] M. FRÉCHET : Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon Sect. A*, 9:53–77, 1951.
- [Fé56] R. FÉRON : Sur les tableaux de corrélation dont les marges sont données, cas de l'espace à trois dimensions. *Publications de l'Institut de Statistique de l'Université de Paris*, 5:3–12, 1956.
- [Gaf13] Félix GAFFIOT : Dictionnaire latin français (sur site Web dicFro). <http://www.micmap.org/dicfro/search/gaffiot/copula>, cité en Juillet 2013.
- [Gum61] E. J. GUMBEL : Bivariate logistic distributions. *Journal of the American Statistical Association*, 56:335–349, 1961.
- [Hoe40] Wassily Hoeffding : Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, 5(3):179–233, 1940. Réimprimé sous *Scale-invariant correlation theory*. Fisher NI, Sen PK (eds) The Collected Works of Wassily Hoeffding. Springer, New York pages 57-107.
- [Hoe41] Wassily Hoeffding : Masstabinvariante Korrelationsmasse für diskontinuierliche Verteilungen. *Arkiv für matematischen Wirschaften und Sozialforschung*, 7:49–70, 1941. Réimprimé sous *Scale-invariant correlation measures for discontinuous distributions*. Fisher NI, Sen PK (eds) The Collected Works of Wassily Hoeffding. Springer, New York pages 109-133.
- [Joe01] Harry JOE : *Multivariate Models and Dependence Concepts*, volume 73 de *Monographs on statistics and applied probability*. Boca Raton : Chapman & Hall/CRC, 2001. (publié à l'origine : London ; New York : Chapman & Hall ; First edition, 1997).
- [Mui82] Robb J. MUIRHEAD : *Aspects of Multivariate Statistical Theory*. Wiley series in probability and mathematical statistics. John Wiley & Sons, Inc., 1982.
- [Nel06] Roger B. NELSEN : *An Introduction to Copulas*. Springer Series in Statistics. Springer Science+Business Media, Inc., New York, 2e édition, 2006.
- [Skl59] Abe SKLAR : Fonctions de répartition à  $n$  dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231, 1959.

MYRIAM CHABOT

DÉPARTEMENT DE MATHÉMATIQUES, UNIVERSITÉ DE SHERBROOKE

Courriel: Myriam.Chabot@USherbrooke.ca